DOCUMENT RESUME

ED 191 902 TM 800 526

AUTHOR Hetzel, Donna C.; And Others
TITLE A Quantitative Synthesis of the Effects of Open Education.
PUB DATE Apr 80
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Conventional Instruction; Elementary Secondary Education; *Institutional Characteristics; Open Education; *Outcomes of Education; *Research Design; *Student Characteristics
IDENTIFIERS *Meta Analysis

ABSTRACT

By employing meta-analysis procedures to build on the reviews of Horwitz and Peterson, specific effects of open education are quantified and related to school and student characteristics. From the 45 studies located in ERIC, the Education Index, and Horwitz's monograph, student, setting, and research design characteristics were coded. Twenty-three affective and cognitive outcomes including subject-matter achievement, self-concept and attitude toward teacher were examined. The broad range of outcomes represents the diversity of effects to be expected from a school-level intervention. Overall, open education has slightly higher outcomes than does traditional education, but the difference is not significant. The study-weighted outcome t-test shows an average open education superiority of .13 traditional-group standard deviations. Results reinforce the view that open education offers an eminently viable option to educators in search of alternatives to traditional teaching methods. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191902

A Quantitative Synthesis of the Effects of Open Education

Donna C. Hetzel, Sue Pinzur Rasher,
Larry Butcher and Herbert J. Walberg
University of Illinois at Chicago Circle

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. Rasher

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Effects of Open Education

A paper to be presented at the 1980 Annual Meeting of the American Educational Research Association, Boston, Massachusetts

TM 800526

A QUANTITATIVE SYNTHESIS OF THE EFFECTS OF OPEN EDUCATION

In the late 1960's and early 70's American teachers were being encouraged by influential educational writers to adopt teaching practices which had come to be associated with what was termed "Open Education". These first writings were mainly descriptive, sometimes philosophical and occasionally instructive. By the mid-70's, the impact of Open Education was being felt throughout the United States. At the same time, there was an increasing emphasis on "accountability" in the country's education community. This combination led to an outpouring of research dealing with the efficacy of open vs. traditional education. In 1976, Horwitz was able to cite over two hundred studies in his, the first comprehensive, review of this research.

In an attempt to establish statistically sound methods of dealing with rapidly expanding research literature, Gene Glass (1976) and others have proposed procedures for research in "meta-analyses" that are attempts to make sense of findings which are "fragile;..(which) vary in confusing irregularity across contexts, classes of subjects, and countless other factors." (p. 3)

This quantitative synthesis of the effects of open education builds on the reviews and integrative attempts of Horwitz (1976, 1979) and Peterson (1979) by employing meta-analyses proposed in the works of Gage, Glass, Walberg and others.

PRIOR SYNTHESES

The first compilation and anlysis of the comparative research literature on open and traditional education was carried out by Horwitz (1976) and appeared as a monograph of the North Dakota Study Group on Evaluation. In his initial attempt to integrate these studies, Horwitz used a "box score" or "voting" method. Each study was assigned to one of four categories based on its published conclusions, i.e., 1) significantly favoring open education, 2) significantly favoring traditional teaching, 3) no significant differences, and 4) mixed results. After each of his over two hundred entries had been assigned, he tallied the number of studies within each category, drawing conclusions and determining "winners" on the basis of plurality. Light and Smith (1971) suggest that this "winning" category gives the best estimate of the direction of the relationship between the independent and dependent variable, but remain skeptical of its utility as contrasted with a re-analysis of combined data from studies.

Peterson (1979) pointed out that the "voting" method ignored "good descriptive information in the studies" (p. 60) and calculated Glass' (1976) effect sizes. By subtracting the mean on the outcome variable of the control group from the mean on the same outcome variable of the experimental group and dividing the result by the standard deviation of the control group, one can calculate the strength or importance of the effect rather than

only the direction. (Glass, 1976, p. 6) Peterson calculated effect sizes for both those studies cited by Horwitz which reported sufficiently detailed information, and for twenty additional studies which she had found. In total, she considered forty-five studies and calculated mean and median effect sizes for twelve educational outcomes, five cognitive and seven affective. Peterson concluded that "although a more direct or traditional teaching approach may be slightly better, on the average, than an open approach for increasing students' (cognitive) achievement, an open approach appears to be better than a more direct approach for increasing students' creativity, independence, curiosity, and favorable attitudes toward school and learning." (p. 67) In addition, she reports that the "research suggests that some kinds of students may do better in an open approach and others may do better in a more direct approach." (p. 67)

It is to this issue as well as the general comparison of open education and traditional education that this research synthesis was especially addressed. That is, this study focuses on questions such as: "Which approach is more successful for students at different grade levels, of different abilities, at different socioeconomic levels, etc.?" as well as queries regarding the relationship between open or traditional approaches and achievement in specific subject matter areas and selected psychological behaviors.

METHODOLOGY

Data Collection:

In our review of research literature we used three main sources: ERIC, The Education Index and the Horwitz monograph (1976). Several bibliographies and lists of references were used to insure completeness of the search. Because of the ambiguity surrounding the definition of Open education (see Horwitz, 1979 for a fuller discussion of this issue) we felt obliged to use fourteen descriptors when doing the ERIC search. (See Appendix A) The Education Index, on the other hand, had but one all-inclusive category: "Open Plan Schools". Only research carried out in North America was considered. Unpublished dissertations were not included.

Data Coding:

The second phase of the work required us to look at the variables and the outcome measures which had been considered in the research. There were basically three sets of independent variables. One set dealt with the nature of the student: sex grade, race, income, ability; another, the nature of the setting: operational definition, school type, sample size, internal or external examiner; and the third, the nature and rigor of the research design: unit of analysis, pretest/post-test, general bias, selection bias, contamination, mortality, representative sample, controls. (See Appendix B for definitions of these

variables) Also, when possible, reliability, level of significance, effect size and direction were recorded and/or calculated.

Twenty-three outcome types were considered by the research. (Appendix C lists both the measures and their frequencies) Data culled from each study was initially recorded on specially designed coding sheets (See Appendix D) and later rendered to data cards.

Data Analyses:

A number of alternative analyses, both effect size and sign and multivariate and univariate, both comparison-weighted and study-weighted, were carried out to explore the overall contrast of open and traditional education as well as specific contrasts of the two on different types of students and outcomes, under alternative operational definitions of open education, and in studies evidencing different threats to methodological validity. It is important to recognize that the comparisons of effect sizes within studies cannot be considered statistically-independent pieces of evidence since what may enhance cognitive learning may also improve self-concept, and what helps boys may also help girls, etc. Taking an overall average of comparisons, moreover, without weighting them by study, weights each study in direct proportion to the number of detailed statistics reported in the study, e.g., by grade level, sex and socioeconomic groups rather than weighting each study equally or according to its

methodological or operational validity. Nevertheless, both comparison-weighted and study-weighted effect sizes and signs are reported partly out of methodological interest in contrasting the two and partly because the comparison-weighted effect sizes serve as intermediate calculations before analyses of variance. Analyses of variance test specific interaction effects but weight each study equally and independently without regard for different categories of operational and methodological validity. Even though effect sizes are more powerful in detecting systematic differences across studies because they indicate magnitude as well as direction of differences, as stated, some studies report insufficient statistics for the calculation of effect sizes. Most of these studies, however, do report signs, that is, which group is higher on the outcome measure and, for this reason, all the available results for studies are analyzed by sign as well as effect size.

RESULTS

Table 1 shows that the search and selection procedures yielded 580 comparison-weighted effect sizes from twenty-five separate studies. A positive effect size, of course, indicates that the open education group scored higher on the outcome measure than did the traditional group. The data are fairly normally distributed, with few outlying cases. The mean comparison-weighted effect size is .016, and the study-weighted mean is .134; these figures place the open education groups

at approximately 50.5 and 56 percentiles of the traditional group norm. Neither difference is significantly different from zero at even the .10 level as indicated by the T-tests of .846 and 1.199, respectively. Thus, while open education appears to have very slightly higher outcomes on average and in general than does traditional education, the difference is not significant and is likely to have occurred by chance.

Table 2 shows the frequency of positive and negative signs and the mean effect sizes by outcome type. The last row in the table shows that 54.9 per cent of the 869 comparison-weighted signs and 53.0 per cent of the 198 study-weighted signs are positive, which confirms the slight preponderance of results favoring open education; neither of these percentages however, is significantly different from a 50-50 per cent split in a binomial probability test. (Dixon & Massey, 1969) The comparison-weighted T-test for all effect sizes is non-significant; but, the study-outcome-weighted T-test shows an average open education superiority of .13 traditional-group standard deviations. The only specific outcome that is significant by itself is self-concept. All the results mentioned in this paragraph must be largely discounted because they violate the multivariate, independence, and degrees-of-freedom assumptions; for example, the one specific outcome, self-concept, that is significant at the .01 level is what would be expected from comparing two groups on approximately twenty outcomes; even the 132 study-weighted effect sizes, moreover, weight each study proportionally to the

number of types of outcomes represented in the study, violating both the weighting and independence assumptions.

The results presented in Table 2 are comparable to the studies by Horwitz (1976) and Peterson (1979). In Horwitz' second study he examined 362 comparison-weighted signs from approximately 200 studies by tallying the number of studies favoring open education, traditional teaching, no significant differences, or mixed (ambiguous) results. He found little or no difference between open and traditional education on academic achievement. Self-concept, creativity, curiosity, adjustment, anxiety, and locus of control were more positive outcomes for open education. Also, positive attitude toward school, independence, and cooperation led him to conclude that open education sometimes has measurable advantages, sometimes makes no difference, and rarely does harm.

Peterson calculated effect sizes from twenty-five studies included in Horwitz 1976 analysis and twenty additional studies. She found achievement outcomes favor open education by .1 standard deviation. Open education students performed slightly lower on achievement tests but slightly higher on tests of more divergent thinking, such as creativity and problem solving. On affective outcomes, as stated, Peterson found slightly positive results of open education on attitude toward school, attitude toward teacher, independence and curiosity, and little or no difference on self-concept, locus of control, and anxiety.

Table 3 shows the analyses of variance and T-tests for all coded characteristics of the operational definition, sample characteristics, and methodological features of the studies. Although many of the comparison-weighted analyses' F- and T-tests are significant, none of the study-weighted effect sizes are significantly different across definition, sample, or methodological characteristics. There are slight, non-significant tendencies for study-weighted effect sizes to be higher under the Walberg-Thomas and nominal definitions than under the Weiss-Traub definition and higher for studies with individual students rather than classes or schools as units of analysis. The results of all the analyses, however, are most reasonably interpreted as showing a very slight and non-significant advantage of open education overall that does not differ across sample, operational, and methodological characteristics of the studies. Because of the non-differences across characteristics, interactions of student and sample variables with operational characteristics were not calculated.

DISCUSSION

In contrast to other quantitative research syntheses using similar methodologies that have been carried out on reviews of research of school learning, (Walberg, Schiller, & Haertel, 1979) class size (Glass & Smith, 1979), and student motivation (Uguroglu & Walberg, 1979), the present syntheses show notable inconsistencies. To say that the outcome measures

are insensitive seems untenable in that in many cases they reflect the expressed goals of open education (Spodek & Walberg, 1975) and have proven sensitive to educational treatments and conditions in the syntheses mentioned above.

One possible explanation is the lack of operational validity in the construct, open education. Walberg and Thomas (1974), however, carried out an elaborate content analysis of open education writings, carefully drafted items on the basis of the analysis, validated them with expert practitioners, and showed the inter-states reliability and discriminant validity in the United States and England. Studies using the Walberg-Thomas scales show results similar to architectural and nominal definitions. Weiss-Traub's defined open education, on the other hand, does less well on average, but again, without significance. Future research might exploit the full metric precision of the open education operational measures as well as control variables such as IQ and pretests by regressing outcomes on them as continuous rather than as dichotomous variables.

It is also clear from Table 3 that most of the studies of open education are seriously flawed in several methodological ways, any one of which can vitiate the creditability of the findings. Future studies should be designed so as to remove the Cook-Campbell (1976) threats to methodological validity. As noted above, however, many studies of other educational treatments are equally flawed and yet produce consistently positive

results (Walberg, Schiller, and Haertel, 1979). Thus it is difficult to imagine that a great number of additional well-designed studies will show results much different than revealed by the present comprehensive sample of published studies.

We would conclude, as did Horwitz (1979) that "political decisions to support or not support open education will ... continue to be made regardless of research evidence" (p. 86). We feel though, that this study further reinforces the view that open education offers an eminently viable option to educators who are in search of alternatives to traditional teaching methods.

References

Cook, T. D. & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), The handbook of industrial and organizational psychology. Rand McNally, 1976.

Dixon, W. J. & Massey, F. J. Introduction to statistical analysis. New York: McGraw-Hill, 1969.

Glass, G. V Primary, secondary, and meta-analysis of research, Educational Researcher, 1976, 5 (10), 3-8.

Glass, G. V & Smith, M. L. Meta-Analysis of research on the relationship of class-size and achievement, Evaluation Policy Analysis, 1979, 1, 2-15.

Horwitz, R. A. Psychological effects of open classroom teaching on primary school children: A review of the research, Grand Forks, N. D.: North Dakota Study Group on Evaluation, June 1976.

Horwitz, R.A. Psychological effects of the "open classroom," Review of Educational Research, 1979, 49 (1), 71-86.

Light, R. J. and Smith, P. V. Accumulating evidence: Procedures for resolving contradictions among different research studies, Harvard Educational Review, 1971, 41, 429-471.

Peterson, P. L. Direct instruction reconsidered. In P. J. Peterson & H. J. Walberg (Eds.), Research on teaching: Concepts, findings, & implications. Berkeley, Cal.: McCutchan, 1979.

Spodek, B. & Walberg, H. J. (Eds.), Studies in open education. New York: Agathon, 1975.

References (con't)

Uguroglu, M. E. & Walberg, H. J. Motivation and achievement: A quantitative synthesis, American Educational Research Journal, 1979, 16 (4), 375-389.

Walberg, H.J., Schiller, D. & Haertel, G. D. The quiet revolution in educational research, Phi Delta Kappan, 1979, 61 (3), 179-183.

Walberg,H. J. & Thomas, S. C. Defining open education, Journal of Research and Development in Education, 1974, 8, 4-13.

APPENDIX A

Fourteen Descriptors used in the ERIC search

1. Classroom design
2. Discovery learning
3. Experimental schools
4. Flexible classrooms
5. Flexible facilities
6. Humanistic education
7. Individualized programs
8. Open education
9. Open plan schools
10. Progressive education
11. School space
12. Self-directed classrooms
13. Student-centered curriculum
14. Ungraded classes

APPENDIX B

Four Kinds of Validity in Experiments and Quasi-Experiments

(adapted from Cook and Campbell, 1974)

Kind — General and Some Illustrative Questions

A. Statistical Conclusion: Do the presumed caused and effects covery?

1. Statistical power: Is the sample size large enough to reject the null hypotheses at a given level of probability (or, as preferred here, estimate coefficients with a reasonably small margin of error)?
2. Error rate: Do the "significant" effect that 1 in 20 on the average in which the probability of the evidence, given the null hypothesis, is less than .05.
3. Reliability of measures: Are the measurement instruments internally consistant or stable enough to detect changes during, or differences in, treatments?
4. Reliability of treatment implementation: Have the units in a treatment actually undergone the same conditions?

B. Internal: Are there other plausible explanations of the covariation between causes and effects?

5. Maturation: Have factors within the units, rather than the presumed external causes, brought about changes or differences?
6. History: Have external factors in the environment, other than the presumed cause, brought about the result?
7. Selection bias: Do pre-existing differences among the groups account for later differences on the dependent variable?

APPENDIX B

Four Kinds of Validity in Experiments and Quasi-Experiments

(adapted from Cook and Campbell, 1974)

Kind — General and Some Illustrative Questions

A. Statistical Conclusion: Do the presumed causes and effects cover?

1. Statistical power: Is the sample size large enough to reject the null hypotheses at a given level of probability (or, as preferred here, estimate coefficients with a reasonably small margin of error)?
2. Error rate: Do the "significant" effect that 1 in 20 on the average in which the probability of the evidence, given the null hypothesis, is less than .05.
3. Reliability of measures: Are the measurement instruments internally consistent or stable enough to detect changes during, or differences in, treatments?
4. Reliability of treatment implementation: Have the units in a treatment actually undergone the same conditions?

B. Internal: Are there other plausible explanations of the covariation between causes and effects?

5. Maturation: Have factors within the units, rather than the presumed external causes, brought about changes or differences?
6. History: Have external factors in the environment, other than the presumed cause, brought about the result?
7. Selection bias: Do pre-existing differences among the groups account for later differences on the dependent variable?

8. Contamination and compensation: Do untreated control groups imitate or somehow gain the benefits of a treatment or do they enter into competition and expend extra energies and therefore do better than they otherwise would?

9. Mortality: Do differential drop out rates in the treatment groups account for differences on the dependent variable obtained at the end of the experiment.

C. External: Can the cause-effect be generalized to other times, units, and settings?

10. See earlier sections of these reviews.

D. Construct: Are there other plausible hypotheses about both the cause and effect constructs?

11. Operational irrelevency: Does the treatment as administered contain Hawthorne, experimentor-bias, placebo, or other unitended causes or artifacts that actually produce the effects?

12. Effect and time restrictions: Would the treatment affect the different times thus producing "fade out" and "sleeper" effects?

APPENDIX C

Outcome Measures and Frequencies in the Research Literature

1.	Reading	(100)	13.	Locus of control	(5)
2.	Math computation	(103)	14.	Cooperation	(4)
3.	Math problem solving	(80)	15.	General achievement	(56)
4.	Language arts	(122)	16.	General affective gain	(7)
5.	Spelling	(51)	17.	Science	(12)
6.	Vocabulary	(37)	18.	Social Studies	(15)
7.	Word discrimination	(13)	19.	Music	(1)
8.	Self concept	(66)	20.	Art	(1)
9.	Concept of school	(90)	21.	Attitude toward teacher	(28)
10.	Creativity	(13)	22.	Motivation	(4)
11.	Independence	(12)	23.	Other	(25)
12.	Adjustment, security	(10)			

APPENDIX D

Coding Sheet for data collection

Identification: (20)

Author, ERIC No.

Operational definition: (1)

1= other than listed here
2= architectural
3= nominal
4= home made instrument
5= Weiss-Traub
6= Barth Scale
7= Walberg-Thomas

Sex: (1)

1= boys
2= mixed
3= girls

Grade: (2)

1-12= grades
13= pre-school
14= kindergarten
15= high school

Race: (1)

0= not specified
1= all black
2= all white
3= all other
4= mixed

Income: (1)

0= not specified
1= all low
2= all high
3= all middle
4= mixed/varied

Ability: (1)

0= not specified
1= all low
2= all high
3= all middle
4= mixed/varied

School type: (1)

0= not specified
1= public
2= private
3= mixed

Sample size: (4)

0= not specified
4 spaces for exact N

Unit of analysis: (1)

0= not specified
1= individual
2= class/grade
3= pod, module, etc.
4= school
5= varied

Pre-post: (1)

1= no
2= not specified
3= yes

General Bias: (1)

1= biased
2= not observable or specified
3= non-biased

Selection Bias: (1)

1= biased
2= not observable or specified
3= non-biased

Contamination: (1)

1= evidenced
2= non-indicated
3= no evidence

Differential mortality: (1)

1= evidence
2= non-indicated
3= no evidence

Coding Sheet for data collection

Representative Sample: (1)

1= non-random
2= not indicated
3= random sample

Examiner: (1)

1= inside
2= not indicated
3= outside

Controls: (1)

0= no evidence of control
1= control for IQ or SES

Outcome type: (2)

1= reading
2= math computation
3= math problem solving
4= language arts
5= spelling
6= vocabulary
7= word discrimination
8= self-concept
9= school-concept (attitude twd. school/teacher)
10= creativity
11= curiosity
12= independence
13= adjustment, security
14= locus of control
15= cooperation
16= general acheivement
17= general affective gain
18= science
19= social studies
20= music
21= art
22= persistence/motivation
23= other

Reliability: (3)

0= not specified
level _____ (e.g., .67)
1 2 3

Effect Size: (6)

0= not specified
size: _____ (e.g., 1.672)
1 2 3 4 5 6

calculated:

$$\frac{\bar{X}_{exp.} - \bar{X}_{con.}}{SD_{con.}}$$

Level of Significance: (4)

0= not specified
level: _____ (e.g., .001)
1 2 3 4

Direction: (1)

1= traditional better
2= no difference
3= open better

Identification	
Operational Def.	21
Sex	22
Grade	23
Race	25
Income	26
Ability	27
School type	28
Sample size	29
Unit of analysis	33
Pre-post	34
General bias	35
Selection bias	36
Contamination	37
Mortality	38
Representative Sample	39
Examiner	40
Controls	41
Outcome type	42
Reliability	44
Effect size	47
Significance	53
Direction	57

Table 1

Frequency Distributions and Univariate Statistics of Effect Sizes

	Comparison-Weighted Effect Sizes		Study-Weighted Effect Sizes	
	Frequency	Percent	Frequency	Percent
3.0 - 3.4	1	0.2	0	--
2.5 - 2.9	0	--	0	--
2.0 - 2.4	0	--	0	--
1.5 - 1.9	0	--	0	--
1.0 - 1.4	16	2.7	1	4.0
0.5 - 0.9	52	9.1	2	8.0
0.0 - 0.4	258	44.5	14	56.0
- 0.0 - - 0.4	182	31.4	6	24.0
- 0.5 - - 0.9	53	9.1	2	8.0
- 1.0 - - 1.4	14	2.4	0	--
- 1.5 - - 1.9	3	0.5	0	--
- 2.0 - - 2.4	1	0.2	0	--
Total	580	100.1	25	100.0
$\bar{X}$ (S.D.)	0.016	(0.457)	0.134	(0.559)
T	0.846		1.199	

Note: Neither T-test is significant at the .10 level.

Table 2

Signs and Effect Sizes by Weighting and Outcome Type

Outcome Type	Comparison-Weighted Results						Study-Weighted Results					
	Signs		Effect Sizes				Signs		Effect Sizes			
	Number	Percent Positive	Number	$\bar{X}$	S.D.	T	Number	Percent Positive	Number	$\bar{X}$	S.D.	T
General Achievement	55	58.2	17	0.28	0.46	2.49*	10	70.0	5	0.19	0.29	1.46
Reading and Language Arts												
Language Arts	121	44.6	91	-0.06	0.36	-1.51	19	36.8	14	-0.02	0.53	-0.14
Reading	99	63.6	69	0.05	0.35	1.29	24	58.3	14	0.25	0.78	1.20
Spelling	50	38.0	45	-0.11	0.32	-2.28*	13	30.8	10	-0.14	1.92	-0.23
Vocabulary	56	51.8	42	0.06	0.29	1.28	11	45.5	8	0.05	0.00	0.00
Word Discrimination	13	46.2	9	-0.18	0.37	-1.44	6	33.3	5	-0.28	0.00	0.00
Other Cognitive Measures												
Art	1	100.0	1	0.85	0.00	--	1	100.0	1	0.85	0.00	0.00
Math Computation	103	50.5	75	-0.06	0.47	-1.16	21	42.9	14	-0.17	0.75	0.85
Math Problem Solving	80	52.5	70	-0.01	0.45	-0.15	16	43.7	12	0.06	0.52	0.40
Music	1	0.0	1	-0.24	0.00	--	1	0.0	1	-0.24	0.00	0.00
Science	12	33.3	11	-0.34	0.65	-1.76	6	33.3	6	-0.48	0.00	0.00
Social Studies	14	35.7	11	-0.32	0.68	-1.56	5	20.0	5	-0.30	0.00	0.00

Table 2 (con't)

Outcome Type	Comparison-Weighted Results						Study-Weighted Results					
	Signs		Effect Sizes				Signs		Effect Sizes			
	Number	Percent Positive	Number	$\bar{X}$	S.D.	T	Number	Percent Positive	Number	$\bar{X}$	S.D.	T
Affective Measures												
General Affective Gain	7	57.1	0	--	--	--	2	50.0	0	--	--	--
Adjustment Anxiety, Security	10	50.5	6	0.06	0.49	0.30	3	33.3	2	0.10	0.00	0.00
Attitude Toward Teacher	28	85.7	0	--	--	--	2	100.0	0	--	--	--
Cooperation	4	25.0	0	--	--	--	2	50.0	0	--	--	--
Creativity	13	76.9	12	0.17	0.41	1.43	5	80.0	4	0.12	3.06	0.08
Independence	12	66.7	5	0.10	0.45	0.48	4	100.0	2	0.19	0.00	0.00
Locus of Control	5	100.0	5	0.39	0.37	2.32	3	100.0	3	0.62	0.84	1.28
Motivation	4	25.0	4	-0.10	0.23	-0.85	1	0.0	1	-0.10	0.00	0.00
School-Concept	90	62.2	44	0.09	0.48	1.38	20	75.0	10	0.14	0.59	0.75
Self-Concept	66	63.6	43	0.22	0.46	3.11**	17	70.6	10	0.54	0.68	2.51*
Other	25	56.0	19	0.19	0.94	0.88	6	50.0	5	-0.26	0.00	0.00
Total	869	54.9	580	0.01	0.45	0.54	198	53.0	132	0.13	0.56	2.67***

* $p \leq .05$
** $p \leq .01$

Table 3

ANOVAs

	Comparison Weighted Results							Study-Weighted Results						
	Signs			Effect Sizes				Signs			Effect Sizes			
	% Positive	N	F	$\bar{X}$	S.D.	N	F/T	% Positive	N	F	$\bar{X}$	S.D.	N	F/T
Operational Definition			10.66***				6.07***			0.58				0.35
Architectural	0.69	202		0.07	0.15	133	-5.24***	0.70	6		0.17	0.00	2	----
Nominal	0.47	449		-0.03	0.54	298	-0.86	0.67	25		0.22	0.64	13	1.24
Home made inst.	0.57	128		0.04	0.40	59	0.84	0.62	6		0.04	0.00	2	----
Weiss-Traub	0.21	14		-0.21	0.43	14	1.79	0.25	2		-0.18	0.00	2	----
Walberg-Thomas	0.75	71		0.18	0.29	71	5.29***	0.78	5		0.18	0.29	5	1.41
Other	0.22	9		-0.50	0.92	9	-1.63	0.22	1		-0.50	----	1	----
Sex			5.63**				7.38***			0.04				0.02
Boys	0.62	201		0.06	0.34	146	2.18*	0.62	5		0.10	0.41	3	0.41
Mixed	0.50	480		-0.05	0.51	306	-1.74	0.66	35		0.15	0.60	19	1.06
Girls	0.61	166		0.13	0.44	109	3.21**	0.58	3		0.07	----	1	----
Race			1.99				7.14***			0.00				0.00
Black	0.72	92		0.08	0.15	77	4.77***	0.72	2		0.13	----	1	----
White	0.72	202		0.19	0.40	148	5.73***	0.79	8		0.29	0.44	5	1.50
Other	1.00	1		1.45	----	1	----	----	0		----	----	0	----
Mixed	0.60	110		0.28	0.47	41	3.82***	0.73	7		0.43	0.68	4	1.25

ANOVA'S con't

	Comparison Weighted Results							Study-Weighted Results						
	Signs			Effect Sizes				Signs			Effect Sizes			
	% Positive	N	F	$\bar{X}$	S.D.	N	F/T	% Positive	N	F	$\bar{X}$	S.D.	N	F/T
Grade			2.61**				3.33***			0.00				0.35
Kindergarten	0.75	4		0.15	0.20	4	1.44	----	0		----	----	0	----
Grade 1	0.42	66		0.02	0.72	45	0.18	0.42	1		-0.00	----	1	----
Grade 2	0.61	142		0.06	0.42	44	0.95	0.71	5		0.18	0.42	2	0.59
Grade 3	0.61	191		0.12	0.34	126	3.92***	0.55	4		0.16	0.61	4	0.53
Grade 4	0.48	83		-0.20	0.61	59	-2.58*	0.71	7		-0.05	0.82	2	0.09
Grade 5	0.51	143		-0.08	0.41	101	-2.04*	0.66	10		-0.09	0.57	5	-0.35
Grade 6	0.55	95		0.05	0.47	79	1.04	0.50	4		0.09	0.53	2	0.24
Grade 7	0.40	42		0.00	0.53	29	0.02	0.41	3		0.58	0.95	2	0.87
Grade 8	0.59	58		0.02	0.27	57	0.53	0.82	1		0.44	----	1	----
Grade 9	0.00	4		----	----	0	----	----	0		----	----	0	----
Grade 12	1.00	1		----	----	0	----	----	0		----	----	0	----
Varied Primary	0.82	22		0.19	0.19	22	4.69***	0.82	1		0.19	----	1	----
Income			5.82***				5.24**			0.13				0.76
Low	0.67	85		0.32	0.59	73	4.61***	0.67	5		0.58	0.84	3	1.19
High	0.89	28		0.28	0.31	26	4.62***	0.92	1		0.28	----	1	----
Middle	0.57	141		0.03	0.39	69	0.59	0.75	12		0.15	0.40	7	0.99
Mixed	0.48	71		0.67	0.81	2	1.17	0.68	4		0.67	----	1	----

ANOVA'S con't

	Comparison Weighted Results							Study-Weighted Results						
	Signs			Effect Sizes				Signs			Effect Sizes			
	% Positive	N	F	$\bar{X}$	S.D.	N	F/T	% Positive	N	F	$\bar{X}$	S.D.	N	F/T
Ability			7.47***				2.64*			0.81				
Low	0.71	24		0.44	0.37	7	3.18*	0.74	3		0.48	0.00	2	----
High	0.20	30		-0.17	0.38	12	-1.54	0.37	2		-0.08	0.00	2	----
Middle	1.00	6		0.52	0.63	3	1.44	1.00	3		0.44	0.65	2	0.96
Mixed/varied	0.53	303		-0.02	0.59	167	-0.55	0.67	10		0.27	0.82	5	0.73
School Type			2.47				0.62	Fewer than 2 non-empty groups						0.06
Public	0.55	867		0.01	0.46	579	0.68				0.12	0.58	22	1.00
Mixed	1.00	3		0.27	0.08	2	4.62***				0.27	----	1	----
Sample Size			3.32**				0.57			0.63				0.53
6-50	0.54	301		0.04	0.53	180	1.09	0.50	10		-0.08	0.64	6	-0.31
51-100	0.61	160		0.05	0.51	138	1.18	0.80	10		0.25	0.54	8	1.32
101-200	0.56	71		0.14	0.36	34	2.23*	0.68	8		0.31	0.59	3	0.92
201-1000	0.55	66		0.02	0.30	48	0.46	0.77	7		0.25	0.64	3	0.69
1001 +	0.72	126		0.09	0.15	109	5.86***	0.56	1		----	----	0	----
Unit of Analysis			11.19***				12.13***			0.89				0.48
Individual	0.61	644		0.07	0.42	461	3.73***	0.71	30		0.23	0.56	15	1.59
Class	0.40	43		-0.29	0.72	19	-1.75	0.44	4		-0.16	0.95	2	-0.24
Pod, Module, etc.	0.55	51		0.21	----	1	----	0.68	3		0.21	----	1	----
School	0.33	131		-0.18	0.51	103	-3.62***	0.44	6		-0.02	0.52	6	-0.09
Varied	0.00	3		----	----	0		----	0		----	----	0	----

ANOVA'S con't

	Comparison Weighted Results							Study-Weighted Results						
	Signs		Effect Sizes					Signs		Effect Sizes				
	% Positive	N	F	$\bar{X}$	S.D.	N	F/T	% Positive	N	F	$\bar{X}$	S.D.	N	F/T
Pre-Post			7.82**				0.14			0.14				0.00
No	0.61	413		0.06	0.36	223	2.37*	0.65	25		0.13	0.64	10	0.65
Yes	0.51	338		0.04	0.49	252	1.35	0.59	13		0.12	0.48	8	0.70
General bias			0.00				1.09			0.00				0.08
Biased	0.56	150		0.05	0.43	80	0.94	0.62	3		0.11	0.00	2	----
Non-biased	0.56	73		-0.12	0.35	8	-0.96	0.64	3		0.12	----	1	----
Selection bias			31.67***				24.02***			1.93				1.63
Biased	0.91	32		0.31	0.35	28	4.66***	0.96	3		0.48	0.00	2	----
Non-biased	0.40	218		-0.17	0.50	141	-4.16***	0.51	8		-0.10	0.49	5	-0.47
Contamination			1.09				5.40*			0.43				0.77
Evidenced	1.00	2		0.67	0.81	2	1.17	1.00	1		0.67	----	1	----
No Evidence	0.64	112		0.08	0.34	57	1.78	0.64	5		0.10	0.37	3	0.46
Differential Mortality			9.11**							1.89				
Evidenced	1.00	7	Fewer than 2 non-empty groups					1.00	2	Fewer than 2 non-empty groups				
No evidence	0.43	222						0.47	7					
Representative Sample			14.85***				4.82*			0.02				0.12
Non-random	0.45	418		-0.09	0.48	272	-3.12**	0.67	24		0.18	0.56	13	1.14
Random	0.67	88		0.17	0.56	18	1.28	0.69	8		0.30	0.67	3	0.79

ANOVA'S con't

	Comparison Weighted Results							Study-Weighted Results						
	Signs			Effect Sizes				Signs			Effect Sizes			
	Positive	N	F	$\bar{X}$	S.D.	N	F/T	Positive	N	F	$\bar{X}$	S.D.	N	F/T
Examiner			35.71***				42.57***			0.38				1.48
Inside	0.42	340		-0.13	0.47	241	4.35***	0.59	12		-0.08	0.46	1	-0.40
Outside	0.63	491		0.12	0.42	301	4.88***	0.69	28		0.23	0.59	19	1.49
Controls			26.60***				6.45*			0.07				0.15
No evidence	0.46	429		-0.03	0.51	307	0.99	0.62	16		0.08	0.50	11	0.55
IQ or SES control	0.55	444		0.07	0.38	277	2.93**	0.66	28		0.17	0.62	14	1.05

Note: For the particular computer program used, t's for effect sizes with 0 or near 0 standard deviations could not be calculated (Nie, et.al, 1975).